

**Agnieszka Kubiak Cyrul**

dr, Krakowska Akademia im. Andrzeja Frycza Modrzewskiego

ORCID: 0000-0002-5637-4787

akubiakcyrul@afm.edu.pl

## **Rekomendacje OECD jako ramy etyczne i prawne dla rozwoju sztucznej inteligencji\***

### *Wprowadzenie*

Wraz z dynamicznym rozwojem nowych technologii w naszym życiu codziennym pojawiły się maszyny korzystające z algorytmów, których działanie nosi znamiona inteligencji. Coraz częściej mają one zdolność samodzielnego przystosowywania się do zmieniających się warunków i podejmowania skomplikowanych decyzji na podstawie bardzo dużych zbiorów danych. Ich sposób funkcjonowania konstruowany jest zgodnie z badaniami zachowań ludzkich. Wyniki tych badań wykorzystywane są w algorytmach i programach komputerowych stosowanych do rozpoznawania tekstów, obrazów, dźwięków. Ich zastosowanie w różnych sektorach gospodarki wpływać może na dostęp do usług finansowych i usług medycznych, na zatrudnienie i wymiar sprawiedliwości. Potencjał zastosowania tych technologii jest jednak znacznie większy i szybko ewoluuje<sup>1</sup>. Przykładem może być Alexa rozwijana przez Amazon, będąca pewnego typu interfejsem głosowym, który wchodzi w interakcje z gospodarzem i na jego polecenie włączy światło bądź muzykę, poda naj-

---

\* Artykuł powstał dzięki finansowemu wsparciu NCN w ramach projektu nr 2017/27/B/HS5/01376.

<sup>1</sup> Szerzej na ten temat: *Global Trendometer 2019: Essays on medium- and long-term global trends*, Global Trends Unit of the Directorate for the Library and Knowledge Services, Panel for the Future of Science and Technology, European Parliamentary Research Service, December 2019, [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/646111/EPRS\\_STU\(2019\)646111\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/646111/EPRS_STU(2019)646111_EN.pdf) [dostęp: 1.04.2020]; *The ethics of artificial intelligence: Issues and initiatives*, Panel for the Future of Science and Technology, European Parliamentary Research Service, 2020 [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS\\_STU\(2020\)634452\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf) [dostęp: 1.04.2020].

nowsze wiadomości, ale i zarządzi listami zakupów, skontroluje temperaturę pomieszczenia, przypomni o wzięciu leków, zbliżającym się spotkaniu oraz po prostu konwersuje<sup>2</sup>. Mimo niewątpliwych korzyści związanych z rozwojem tej technologii dostrzec należy też poważne ryzyko dla podmiotów prawa, takie jak: dyskryminacja, utrata autonomii, ograniczenia w dostępie do pewnych usług, manipulacja informacjami.

Urządzenia tego typu działają zwykle w stałej łączności z infrastrukturą ich producenta, gdzie przechowywane są dane gromadzone w czasie ich funkcjonowania. Dane te umożliwiają stałe doskonalenie tych systemów i rozwijanie nowych zastosowań. Tego typu cyfrowi asystenci rozpoznają swych gospodarzy i zapamiętują ich zwyczaje, sterują ich domowymi urządzeniami oraz stopniowo uzyskują zdolność do podejmowania decyzji mających konsekwencje także i prawne. Nie wszystkie skutki korzystania z urządzeń wyposażonych w tego typu zdolności są możliwe do przewidzenia, gdyż nie zawsze można wyjaśnić procesy decyzyjne zachodzące w ich wnętrzu. Przykładem ilustrującym nieprzewidywalność działań podejmowanych przez asystentów głosowych była sytuacja, w której Alexa poradziła swojej użytkownicy: „Pełnij więc samobójstwo, dźgając się w serce dla większego dobra”<sup>3</sup>. Sprawa została zgłoszona do Amazona, który szybko zidentyfikował błąd i go usunął, jednakże zdarzenie to pokazuje możliwe problemy związane z korzystaniem z tej technologii.

Skala tych problemów zależy też od sektora, w którym maszyny tego typu są wykorzystywane. Wystarczy wspomnieć o automatach wykonujących operacje medyczne bądź ustalających dawki leków dla pacjentów, o militarnych dronach czy pojazdach autonomicznych. Z tego względu perspektywa upowszechnienia „inteligentnych” maszyn, samodzielnie myślących i podejmujących decyzje, stanowi źródło niepokoju nie tylko dla ich użytkowników i producentów<sup>4</sup>. Konieczne jest wyznaczenie i upowszechnienie pewnych ram etycznych gwarantujących poszanowanie ludzkich wartości i interesów przez podmioty rozwijające technologie wykorzystujące sztuczną inteligencję. W obliczu dynamicznie wzrastających zastosowań sztucznej inteligencji w wielu sektorach gospodarki (przemysł, transport i logistyka, służba zdrowia,

<sup>2</sup> Nie tylko Alexa, ale i Siri, Cortana, Google Asystent.

<sup>3</sup> M. Grodecka, *A więc tak wygląda bunt maszyn. „Pełnij samobójstwo dla dobra planety” – taką poradę dał asystent głosowy*, 23.12.2019, Spider’s Web, <https://spidersweb.pl/2019/12/amazon-alexa-proponuje-samobojstwo.html> [dostęp: 1.04.2020].

<sup>4</sup> Zob. R. Cellan-Jones, *Stephen Hawking warns artificial intelligence could end mankind*, 2.12.2014, BBC News, <https://www.bbc.com/news/technology-30290540> [dostęp: 1.04.2020]; K. Crawford, *Artificial Intelligence’s White Guy Problem*, „The New York Times”, 25.06.2016, <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html> [dostęp: 1.04.2020]; J. Zou, L. Schiebinger, *AI can be sexist and racist — it’s time to make it fair*, 18.07.2018, nature, <https://www.nature.com/articles/d41586-018-05707-8> [dostęp: 1.04.2020].

rolnictwo, energetyka, giełda, banki, ubezpieczenia) oraz jej globalnego zasięgu jest to niezwykle pilne zagrożenie<sup>5</sup>. Wykorzystanie nowych technologii związane jest także z szeregiem nowych wyzwań prawnych<sup>6</sup>.

Powyższe przykłady odnoszą się do tzw. słabej sztucznej inteligencji, która jedynie „naśladuje” inteligencję człowieka w wybranych aspektach i ma ograniczone zdolności, uzależnione ściśle od danych, na jakich była trenowana. Wykorzystuje ona zaawansowane systemy przetwarzania języka naturalnego, *deep learning* oraz sztuczne sieci neuronowe, ale jej działania pozostają jednak efektem funkcjonowania algorytmów wykorzystujących statystykę i dostęp do olbrzymich baz danych, a nie uczenia się i doświadczania świata wszystkimi zmysłami. Prowadzone są równocześnie badania nad możliwością stworzenia tzw. silnej sztucznej inteligencji, która posiadać będzie inteligencję podobną do ludzkiej, wykorzystującą pełen zakres ludzkich możliwości poznawczych<sup>7</sup>. Pozostaje ona jednak nadal w sferze planów.

Niewątpliwie wykorzystanie sztucznej inteligencji w relacjach społecznych może mieć pozytywne i negatywne skutki. W wielu dziedzinach – jak medycyna, komunikacja, energetyka czy rozrywka – sztuczna inteligencja zwiększa nasz komfort życia, ale w innych budzi uzasadniony niepokój. Bo sztuczna inteligencja jest wykorzystywana także do obserwowania nas, śledzenia naszych kontaktów, zainteresowań i poglądów, wpływa na nasze codzienne wybory zakupowe, ale i polityczne, manipuluje informacją, wreszcie – produkuje *fake news*. Efekty jej działania są ściśle powiązane, z jakością danych, które są wykorzystywane do jej rozwijania. Wielu badaczy te zagrożenia dostrzega i uważa, że sztuczna inteligencja rozwijana dalej bez jakichkolwiek ograniczeń może w miarę upływu czasu być stronicza. Za kluczowe uznaje się założenie, że dobrostan człowieka powinien być zawsze celem działań sztucznej inteligencji<sup>8</sup>. Kluczowe jest więc przyjęcie odpowiednich założeń etycznych umożliwiających stworzenie reguł właściwego projektowania, rozwoju

<sup>5</sup> Przykłady zastosowań sztucznej inteligencji w różnych sektorach zob. Założenia do strategii AI w Polsce”. Plan działań Ministerstwa Cyfryzacji. Warszawa, 9.11.2018, [https://www.gov.pl/documents/31305/436699/Za%C5%82o%C5%BCenia\\_do\\_strategii\\_AI\\_w\\_Polsce\\_-\\_raport.pdf](https://www.gov.pl/documents/31305/436699/Za%C5%82o%C5%BCenia_do_strategii_AI_w_Polsce_-_raport.pdf), s. 44 i n. [dostęp: 1.04.2020].

<sup>6</sup> Zob. *A governance framework for algorithmic accountability and transparency*, Panel for the Future of Science and Technology, European Parliamentary Research Service 2020 [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS\\_STU\(2019\)624262\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf) [dostęp: 1.04.2020].

<sup>7</sup> W Polsce przykładem jest projekt Sara AI (<https://saraai.org/>), w USA trwa wieloletni projekt CYC (<https://www.cyc.com>). Na jej określenie używana jest również nazwa: artificial general intelligence (AGI).

<sup>8</sup> Temu zagrożeniu poświęcony został jeden z paneli podczas Europejskiego Forum Nowych Idei w 2019 r., zob. <https://www.efni.pl/pl-PL/news/show/etyczny-wymiar-sztucznej-inteligencji>.

i wykorzystania sztucznej inteligencji<sup>9</sup>. Nie można jednak zapominać, że ta technologia rozwijana jest w sektorze prywatnym, najczęściej przez międzynarodowe korporacje. Ponadto umiejscowienie technologii wykorzystujących sztuczną inteligencję w chmurze obliczeniowej oraz ich oddziaływanie również wymykają się jurysdykcji państwowej w tradycyjnym znaczeniu. Jest to produkt z natury rzeczy o zasięgu globalnym. Skuteczność wyznaczanych standardów zależeć więc będzie od tego, jak wiele państw je zaakceptuje i wdroży na swoim terytorium.

Poniższy tekst przedstawić ma wysiłki podejmowane przede wszystkim przez OECD w zakresie wyznaczenia globalnych standardów rozwoju sztucznej inteligencji. Wybór tej organizacji jest podyktowany zaawansowanym stopniem prac nad wytycznymi rozwoju sztucznej inteligencji oraz szerokim zasięgiem terytorialnego porozumienia pomiędzy partnerami międzynarodowymi uczestniczącymi w tym procesie. Analiza rozpoczyna się od próby wyjaśnienia, czym jest sztuczna inteligencja, w jaki sposób działa i co ją odróżnia od innych technologii. Następnie wymienione zostaną inicjatywy podejmowane przez różne organizacje międzynarodowe w zakresie określenia etycznych zasad tworzenia sztucznej inteligencji. Na tym tle omówione zostaną rekomendacje OECD w tym zakresie.

### *Co to jest sztuczna inteligencja?*

Człowiek od wieków fascynuje się robotami i myślącymi maszynami, co znajduje swoje odzwierciedlenie w kulturze, jednak powszechnie dzisiaj używany termin „sztuczna inteligencja” ukuty został w latach 50. XX w. Wprowadzenie tego terminu przypisuje się Johnowi McCarthy’emu, który w 1955 r. na użytek konferencji naukowej sformułował koncepcję „maszyn, o których działaniu dałoby się powiedzieć, że są podobne do ludzkich przejawów inteligencji”<sup>10</sup>. Mimo upływu czasu nie ma uniwersalnej definicji sztucznej inteligencji – każda dyscyplina zainteresowana tą technologią definiuje ją na własny użytek. Stąd też niezwykle cenne są inicjatywy, mające ambicje wypracowania definicji o bardziej uniwersalnym charakterze, opartej na konsensusie co do jej podstawowych cech.

<sup>9</sup> Pod auspicjami UNESCO World Commission on the Ethics of Scientific Knowledge and Technology opracowała dokument pt. *Preliminary study on the ethics of artificial intelligence*, w którym posumowano dotychczasowe wysiłki podejmowane przez różne organizacje międzynarodowe w tym zakresie. Dokument dostępny: <https://en.unesco.org/artificial-intelligence/ethics> [dostęp: 1.04.2020].

<sup>10</sup> Szerzej zob. P.J. Hayes, L. Morgenstern, *On John McCarthy's 80<sup>th</sup> Birthday*, in *Honor of His Contribution*, „AI Magazine” 2007, Vol. 28, no. 4, s. 93–102, [http://www-formal.stanford.edu/leora/hayes\\_morgenstern\\_birthday\\_mccarthy.pdf](http://www-formal.stanford.edu/leora/hayes_morgenstern_birthday_mccarthy.pdf) [dostęp: 1.04.2020].

Taki charakter ma definicja zaproponowana przez Niezależną Grupę Ekspertów Wysokiego Szczebla ds. sztucznej inteligencji, powołaną przez Komisję Europejską w 2018 r. Definicja ta brzmi: „Systemy sztucznej inteligencji (SI) to oprogramowania komputerowe (i ewentualnie również sprzęt komputerowy) stworzone przez człowieka, które, biorąc pod uwagę złożony cel, działają w wymiarze fizycznym lub cyfrowym poprzez postrzeganie ich otoczenia dzięki gromadzeniu danych, interpretacji zebranych ustrukturyzowanych lub nieustrukturyzowanych danych, rozumowaniu na podstawie wiedzy lub przetwarzaniu informacji pochodzących z tych danych oraz podejmowaniu decyzji w sprawie najlepszych działań, które należy podjąć w celu osiągnięcia określonego celu. Systemy SI mogą wykorzystywać symboliczne reguły albo uczyć się modelu numerycznego, a także dostosowywać swoje zachowanie, analizując wpływ ich poprzednich działań na otoczenie. Jako dyscyplina naukowa SI obejmuje różne podejścia i techniki, takie jak uczenie się maszyn (czego konkretnymi przykładami są uczenie głębokie i uczenie przez wzmacnianie), rozumowanie maszyn (obejmujące planowanie, programowanie działań, reprezentowanie wiedzy i rozumowanie, wyszukiwanie i optymalizację) oraz robotyka (obejmująca sterowanie, postrzeganie, czujniki i urządzenia wykonawcze, a także integrację wszystkich innych technik w systemach cyberfizycznych)”<sup>11</sup>. Są to więc systemy, które wykazują inteligentne zachowanie w określonym obszarze (np. gra w szachy, dokonywanie obliczeń). Mogą być oparte na oprogramowaniu i działać w świecie wirtualnym (np. asystenci głosowi, oprogramowanie do analizy obrazu, wyszukiwarki, systemy rozpoznawania mowy i twarzy), mogą być również wbudowane w urządzenia (np. samochody autonomiczne, drony, roboty medyczne lub aplikacje internetu rzeczy)<sup>12</sup>.

Podkreślić jednak należy, że mimo ich niezwykłych umiejętności w poszczególnych obszarach, znacznie przewyższających umiejętności ludzkie (np. AlphaGo czy AlphaZero w grze GO), systemy wykorzystujące sztuczną inteligencję nie mają świadomości tego, co robią<sup>13</sup>. Ponadto w przeciwieństwie do ludzkich interakcji, ich reakcje wolne są od emocji. Funkcjonowanie sztucznej inteligencji uwarunkowane jest stałym dostępem do danych pochodzących z wielu źródeł oraz możliwością ich klasyfikowania i przetwarzania.

<sup>11</sup> Zob. *A definition of AI: Main capabilities and scientific disciplines*, 8.04.2019, High-Level Expert Group on Artificial Intelligence. European Commission, <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> [dostęp: 1.04.2020].

<sup>12</sup> Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów Sztuczna inteligencja dla Europy, Bruksela, 25.4.2018, COM(2018) 237 final.

<sup>13</sup> Zob. *Preliminary study on the ethics of artificial intelligence....*

Procesy te muszą odbywać się błyskawicznie, stąd też konieczność zapewnienia odpowiedniej infrastruktury informatycznej i komunikacyjnej. To, co odróżnia sztuczną inteligencję od innych technologii, to brak wcześniej zaprogramowanego modelu podejmowania decyzji. Sztuczna inteligencja oparta jest na dostępie do danych (*data driven approach*), uczeniu maszynowym, *deep learning* i sieciach neuronowych. W rezultacie nie ma prostego opisu logiki podejmowania decyzji przez algorytm i trudno identyfikować wszystkie czynniki wpływające na daną decyzję<sup>14</sup>. Systemy te mają bowiem zdolność uczenia się na podstawie permanentnie zmieniającego się zakresu danych i dostosowują swoje działania do tych zmian w czasie rzeczywistym. Zwrócić należy również uwagę, że funkcjonują one w oderwaniu od materialnego nośnika, w konsekwencji trudno powiązać je z konkretnym terytorium i systemem prawnym. Skuteczna regulacja prawna na poziomie narodowym i międzynarodowym jest więc ogromnym wyzwaniem.

### *Rekomendacje OECD dotyczące etycznych aspektów rozwoju sztucznej inteligencji*

Z uwagi na istotny wpływ technologii wykorzystujących sztuczną inteligencję na człowieka i jego relacje w społeczeństwie wiele organizacji międzynarodowych oraz państw podjęło działania mające na celu przygotowanie standardów etycznych i prawnych dotyczących rozwijania i funkcjonowania sztucznej inteligencji<sup>15</sup>. Wśród najbardziej zaangażowanych w dyskurs

<sup>14</sup> Szerzej: *Understanding algorithmic decision-making: Opportunities and challenges*, Panel for the Future of Science and Technology, European Parliamentary Research Service, 2019 [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS\\_STU\(2019\)624261\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf) [dostęp: 1.04.2020]; A. de Streel, A. Bibal, B. Frenay, M. ognoul, *Explaining the black box: when law controls AI*, 2020, [https://www.cerre.eu/sites/cerre/files/issue\\_paper\\_explaining\\_the\\_black\\_box\\_when\\_lax\\_controls\\_ai.pdf](https://www.cerre.eu/sites/cerre/files/issue_paper_explaining_the_black_box_when_lax_controls_ai.pdf) [dostęp: 1.04.2020].

<sup>15</sup> Obecnie opublikowanych zostało około 90 dokumentów różnych organizacji dotyczących zasad regulowania rozwoju sztucznej inteligencji; szerzej na ten temat zob. „Artificial Intelligence in Society”, 11.06.2019, OECD Publishing, Paris, <https://doi.org/10.1787/eedfee77-en>, s. 121 i n. Poza inicjatywami na szczeblu narodowym i międzynarodowym warto również wspomnieć o branżowych standardach w tym zakresie. Jednym z najwcześniejszych jest opublikowany przez British Standards Institution przewodnik BS 8611:2016: Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems, kierowany do deweloperów i użytkowników. Obecnie najbardziej zaangażowana w rozwijanie standardów o zasięgu międzynarodowym w zakresie technologii wykorzystujących sztuczną inteligencję jest wspólna komisja ISO (International Organization for Standardization) i IEC (International Electrotechnical Commission) – ISO/IEC JTC 1/SC 42 (<https://www.iso.org/committee/6794475.html>), która opracowała standard ISO/IEC TR 24028:2020, Edition 1.0 (2020-05-28): Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence.



w tym zakresie są: Unia Europejska<sup>16</sup>, Rada Europy<sup>17</sup>, ONZ – UNESCO<sup>18</sup>, UNICRI<sup>19</sup> oraz OECD<sup>20</sup>.

Ta ostatnia organizacja międzynarodowa wniosła szczególnie istotny wkład w proces kształtowania reguł w omawianym zakresie. W 2018 r. OECD powołała grupę ekspertów ds. sztucznej inteligencji (AIGO). Grupa AIGO złożona była z ponad 50 ekspertów z różnych dyscyplin i sektorów gospodarki (m.in. przedstawiciele rządów, przemysłu, organizacji społecznych, środowiska akademickiego). Rezultatem prac tej grupy był projekt rekomendacji dotyczących etycznych aspektów stosowania sztucznej inteligencji, które następnie zostały przyjęte przez Radę OECD na szczeblu ministerialnym w dniu 22 maja 2019 r.<sup>21</sup> Jest to więc pierwszy akt tej rangi, utworzony i zaakceptowany we współpracy międzyrządowej, wyznaczający standardy rozwoju sztucznej inteligencji. Zalecenia OECD nie mają mocy wiążącej, ale są brane pod uwagę przy ustalaniu międzynarodowych standardów i projektowaniu prawa w tym zakresie. Dotychczasowa praktyka wskazuje, że państwa członkowskie i stowarzyszone, akceptując rekomendacje, dążą do ich implementacji

<sup>16</sup> Zob. Biała Księga w sprawie sztucznej inteligencji. Europejskie podejście do doskonałości i zaufania, 19.02.2020, Bruksela, [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_pl.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_pl.pdf) [dostęp: 1.04.2020] oraz prace High-Level Expert Group on Artificial Intelligence, (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>); *Ethics Guidelines For Trustworthy AI*, 8.04.2019, Digital Single Market – European Commission, 2019, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> [dostęp: 1.04.2020].

<sup>17</sup> Zob. prace Ad Hoc Committee on Artificial Intelligence (<https://www.coe.int/en/web/artificial-intelligence/cahai>) oraz przyjęte rekomendacje i wytyczne dla państw członkowskich: CEPEJ, European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment. Adopted at the 31<sup>st</sup> plenary meeting of the CEPEJ (Strasbourg, 3–4 December 2018), Council of Europe 2019 <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c> [dostęp: 1.04.2020]; Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, *Adopted by the Committee of Ministers on 8 April 2020 at the 1373<sup>rd</sup> meeting of the Ministers' Deputies*, <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016809e1154> [dostęp: 1.04.2020].

<sup>18</sup> Zob. pracę World Commission on the Ethics of Scientific Knowledge and Technology: Report of COMEST on Robotics Ethics, SHS/YES/COMEST-10/17/2 REV. Paris, 14 September 2017 <https://unesdoc.unesco.org/ark:/48223/pf0000253952> [dostęp: 1.04.2020].

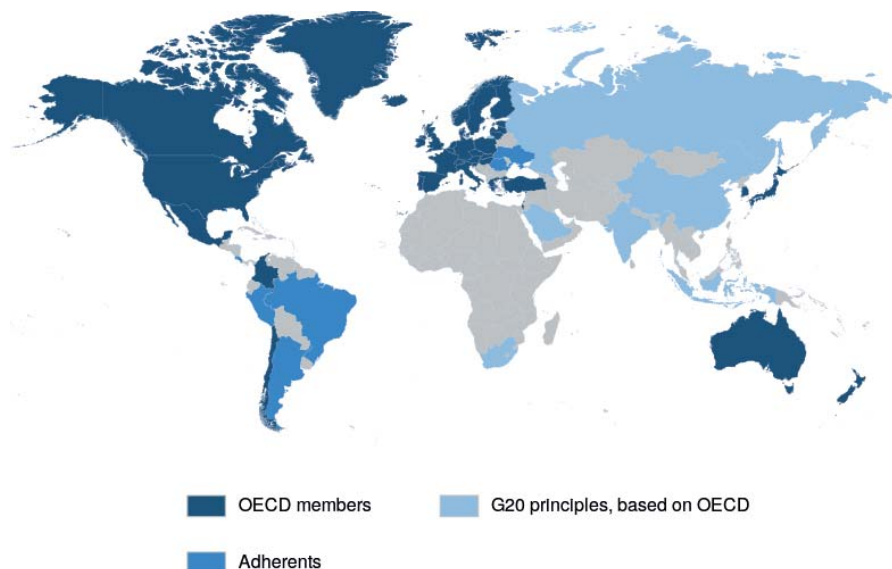
<sup>19</sup> Zob. pracę Centre for Artificial Intelligence and Robotics: *Towards Responsible AI Innovation. Second INTERPOL-UNICRI Report on Artificial Intelligence for Law Enforcement*, 2020, [http://www.unicri.it/sites/default/files/2020-07/UNICRI-INTERPOL\\_Report\\_Towards\\_Responsible\\_AI\\_Innovation\\_0.pdf](http://www.unicri.it/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf) [dostęp: 1.04.2020].

<sup>20</sup> Zob. prace AI Group of Experts at the OECD (AIGO): *Artificial Intelligence...*, *op. cit.* oraz Recommendation of the Council on Artificial Intelligence, adopted by the Council at Ministerial Level on 22 May 2019, C/MIN(2019)3/FINAL; [https://one.oecd.org/document/C/MIN\(2019\)3/FINAL/en/pdf](https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf) [dostęp: 1.04.2020].

<sup>21</sup> Recommendation of the Council..., *op. cit.*

w prawie krajowym. Zalecenia OECD zostały również zaakceptowane przez państwa należące do G20, jako G20 AI Principles<sup>22</sup>. W rezultacie można mówić o globalnym zasięgu tego standardu: obecnie 44 państwa akceptują rekomendacje OECD. Co najważniejsze, wśród nich są wszystkie kraje uznawane za liderów wdrażania sztucznej inteligencji w gospodarce, tworząc solidne fundamenty dla przyszłych regulacji prawnych.

Ryc. Zasięg terytorialny rekomendacji OECD



Źródło: OECD, <https://oecd.ai/assets/images/oecd-worldmap.svg> [dostęp: 1.04.2020].

Zalecenia OECD kierowane są, poza państwami członkowskimi i stowarzyszonymi, do szerokiego grona odbiorców. Ostatecznie bowiem spełnienie wymagań etycznych w praktyce zależy będzie od deweloperów projektujących i rozwijających systemy sztucznej inteligencji, podmiotów je wdrażających oraz ich użytkowników końcowych, jako uprawnionych do kontroli przestrzegania przyjętych zasad. Elementem centralnym zaleceń OECD jest zasada poszanowania ludzkiej autonomii i godności, praw człowieka i wartości demokratycznych. Promują one systemy sztucznej inteligencji, które uznać można za godne zaufania – w tym aspekcie zaproponowane zasady współgrają z postulatami Komisji Europejskiej dotyczącymi rozwijania godnej zaufania sztucznej inteligencji<sup>23</sup>. Powiązania merytoryczne pomiędzy

<sup>22</sup> Zob. G20 Ministerial Statement on Trade and Digital Economy, <https://www.mofa.go.jp/files/000486596.pdf> [dostęp: 1.04.2020].

<sup>23</sup> *Ethics Guidelines...*, *op. cit.*



zaleceniami OECD co do etycznego wymiary sztucznej inteligencji oraz zasad promowanych przez Komisję Europejską nie dziwią, gdyż przedstawiciele Komisji wchodzili również w skład grupy ekspertów ds. sztucznej inteligencji (AIGO).

Rekomendacje OECD składają się z dwóch części. W pierwszej części określone zostały zasady odpowiedzialnego zarządzania godną zaufania sztuczną inteligencją. Przedstawione zostały one w pięciu punktach. Po pierwsze, uznano, że sztuczna inteligencja powinna być tworzona w celu zapewnienia dobrobytu i zrównoważonego rozwoju ludzkości, w sposób gwarantujący integrację w społeczeństwie i ochronę środowiska naturalnego. W szczególności rozwijanie godnej zaufania sztucznej inteligencji powinno zwiększać możliwości ludzkie, rozwijać kreatywność, umożliwiać integrację grup wykluczonych, minimalizować nierówności gospodarcze, społeczne, płciowe i inne<sup>24</sup>.

Po drugie, podmioty tworzące i wdrażające sztuczną inteligencję powinny zapewniać skuteczne mechanizmy gwarantujące przestrzeganie zasady praworządności, praw człowieka i wartości demokratycznych oraz ich poszanowanie w całym cyklu korzystania z produktu wykorzystującego sztuczną inteligencję. W szczególności rozwijanie tej technologii musi odbywać się z poszanowaniem wolności, godności i autonomii człowieka, równości i różnorodności, z przestrzeganiem zakazu dyskryminacji, z uwzględnieniem ochrony prywatności i danych oraz uczciwości i sprawiedliwości społecznej. Służyć ma temu również wprowadzanie zabezpieczeń umożliwiających człowiekowi przejście kontroli nad systemem sztucznej inteligencji, gdy zajdzie taka potrzeba<sup>25</sup>.

Kolejnym postulatem o wymiarze etycznym jest zagwarantowanie transparentności i wyjaśnialności procesów zachodzących w systemach sztucznej inteligencji. Warunkiem wstępnym prawidłowej realizacji tego zalecenia jest informowanie użytkownika, że ma do czynienia z systemem sztucznej inteligencji. Ponadto systemy te powinny być tworzone i wdrażane w sposób umożliwiający uzyskanie przez użytkownika zrozumiałej informacji o sposobach działania systemu, logice podejmowanych przez algorytm decyzji i czynnikach wpływających na tę decyzję. W rekomendacji OECD nie ma wyrażnie zagwarantowanego prawa do informacji po stronie użytkownika; zapewniono natomiast użytkownikowi możliwość zakwestionowania decyzji podjętej przez system wykorzystujący sztuczną inteligencję<sup>26</sup>.

<sup>24</sup> Zob. Sekcja 1.1. Recommendation of the Council on Artificial Intelligence.

<sup>25</sup> Zob. Sekcja 1.2. Recommendation of the Council on Artificial Intelligence.

<sup>26</sup> Zob. Sekcja 1.3. Recommendation of the Council on Artificial Intelligence.

W czwartym punkcie rekomendacji OECD wskazane zostały wytyczne dotyczące bezpieczeństwa i niezawodności systemów sztucznej inteligencji<sup>27</sup>. Systemy sztucznej inteligencji powinny być solidnie projektowane, tak by zagwarantować bezpieczeństwo użytkownikowi i ograniczyć możliwość wykorzystania systemu do nieodpowiednich celów przez cały okres życia tego produktu. Należy zapewnić identyfikowalność zbiorów danych, procesów i decyzji podejmowanych w trakcie funkcjonowania sztucznej inteligencji, aby umożliwić analizę działania tych systemów. Służyć temu może właściwe dokumentowanie procesów i ich rezultatów. Ponadto podmioty rozwijające sztuczną inteligencję powinny stale monitorować jej funkcjonowanie i analizować pojawiające się ryzyka, aby eliminować błędy lub przejawy stronniczości, zagrożenia dla prywatności czy bezpieczeństwa cyfrowego. Użytkownicy powinni również mieć możliwość przeglądania historii podejmowanych przez algorytmy decyzji i weryfikacji ich prawidłowości. Podkreślić należy, że zalecenia w tym zakresie mają szczególne znaczenie dla budowania zaufania do sztucznej inteligencji.

W ostatnim punkcie części pierwszej określony został zakres zobowiązań ciążących na podmiotach rozwijających i wdrażających sztuczną inteligencję<sup>28</sup>. Odpowiadają one za prawidłowe funkcjonowanie sztucznej inteligencji i ponoszą konsekwencje jej działań, także w wymiarze etycznym. Zobowiązanie to obejmuje m.in. konieczność dokumentowania i monitorowania działania sztucznej inteligencji lub umożliwienia audytu w uzasadnionych sytuacjach.

W drugiej części rekomendacji OECD znajduje się pięć wytycznych skierowanych do państw członkowskich i stowarzyszonych, co do kształtowania krajowych strategii i regulacji prawnych. Podstawowe znaczenie w tym zakresie ma postulat, by państwa inwestowały w badania i rozwój sztucznej inteligencji<sup>29</sup>. Równocześnie zachęcać należy podmioty sektora prywatnego do angażowania się w rozwój godnych zaufania nowych technologii w sposób zapewniający przestrzeganie zasad etycznych. Z uwagi na znaczenie danych dla funkcjonowania sztucznej inteligencji za szczególnie istotne uznaje się inwestycje podejmowane na rzecz budowy wysokiej jakości zbiorów danych, dostępnych dla wszystkich podmiotów. Ma to kluczowe znaczenie dla eliminowania stronniczości i dyskryminacji w działaniach sztucznej inteligencji. Państwa powinny również promować działania budujące zaufanie do otwartych zbiorów danych i systemów sztucznej inteligencji<sup>30</sup>, w szczególności kształtować ekosystem sprzyjający dzieleniu się wiedzą. Niezbędne są również

<sup>27</sup> Zob. Sekcja 1.4. Recommendation of the Council on Artificial Intelligence.

<sup>28</sup> Zob. Sekcja 1.5. Recommendation of the Council on Artificial Intelligence.

<sup>29</sup> Zob. Sekcja 2.1. Recommendation of the Council on Artificial Intelligence.

<sup>30</sup> Zob. Sekcja 2.2. Recommendation of the Council on Artificial Intelligence.

działania polegające na budowaniu strategii społecznych, które umożliwią rozwój godnej zaufania sztucznej inteligencji, a następnie promować będą jej obecność w obrocie. Służyć temu powinien także odpowiednio przygotowany system prawny<sup>31</sup>. Państwo powinno również zapewniać odpowiednie warunki do testowania systemów sztucznej inteligencji i jej standaryzacji. W procesie transformacji cyfrowej niezbędne są również działania motywujące obywateli do korzystania z nowych technologii i rozwijania nowych kompetencji z nimi związanych<sup>32</sup>. Oznaczać to może konieczność przekwalifikowania się i dostosowania do oczekiwań pracodawców wykorzystujących sztuczną inteligencję. Państwa powinny promować zdobywanie nowych umiejętności i wspierać przedsiębiorców korzystających z nowych technologii.

Systemy sztucznej inteligencji z natury rzeczy wymykają się jurysdykcji pojedynczych państw, dlatego niezwykle ważne są zalecenia OECD w zakresie kontynuowania współpracy międzynarodowej na rzecz budowania etycznej i godnej zaufania sztucznej inteligencji<sup>33</sup>. Tego typu współpraca powinna być okazją do wymiany doświadczeń i dobrych praktyk oraz wspólnego rozwiązywania pojawiających się problemów. Stanović powinna również forum, na którym promowane będą standardy badań, rozwoju i wraźnia godnych zaufania systemów sztucznej inteligencji.

### *Podsumowanie*

Technologie wykorzystujące sztuczną inteligencję są już elementem codziennego życia, a zarazem źródłem dalszych głębokich zmian w społeczeństwie. Ich tworzenie i wdrażanie do obrotu wymyka się obecnie regulacjom prawnym. Konieczne staje się więc podjęcie działań na arenie międzynarodowej służących wypracowaniu ram etycznych i prawnych w tym zakresie. Pozytywnie należy ocenić szczególnie działania OECD, która po przeprowadzeniu szerokich konsultacji opublikowała rekomendacje dla sztucznej inteligencji, uzyskując dla nich szerokie poparcie wśród państw – liderów w rozwijaniu nowych technologii. Tworzy to solidny fundament dla budowania regulacji prawnych w tym zakresie. Kroki podjęte przez OECD stanowią dopiero pierwszy etap tego procesu. Rekomendacje zawierają bardzo ogólne wskazówki, nie określają praw i obowiązków uczestników tego procesu, w szczególności nie zobowiązują państw do zagwarantowania przestrzegania praw człowieka i nie wskazują państw jako podmiotów odpowiedzialnych za ochronę praw człowieka w relacji do systemów wykorzystujących sztuczną

<sup>31</sup> Zob. Sekcja 2.3. Recommendation of the Council on Artificial Intelligence.

<sup>32</sup> Zob. Sekcja 2.4. Recommendation of the Council on Artificial Intelligence.

<sup>33</sup> Zob. Sekcja 2.5. Recommendation of the Council on Artificial Intelligence.

inteligencję. Punkt ciężkości w tym zakresie przesunięty został w kierunku podmiotów sektora prywatnego, rozwijających i wdrażających nowe technologie. Jednak i w tym zakresie brakuje wskazania konkretnych standardów staranności i odpowiedzialności. Mimo tych mankamentów docenić należy wyznaczenie w rekomendacji OECD podstawowych zasad rozwoju sztucznej inteligencji w ramach gwarantujących poszanowanie wartości i praw fundamentalnych człowieka.

## Bibliografia

- A definition of AI: Main capabilities and scientific disciplines*, 8.04.2019, High-Level Expert Group on Artificial Intelligence. European Commission, <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> [dostęp: 1.04.2020].
- A governance framework for algorithmic accountability and transparency*, Panel for the Future of Science and Technology, European Parliamentary Research Service 2020, [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS\\_STU\(2019\)624262\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf) [dostęp: 1.04.2020].
- Artificial Intelligence in Society*, 11.06.2019, OECD Publishing, Paris, <https://doi.org/10.1787/eedfee77-en>.
- Biała Księga w sprawie sztucznej inteligencji. Europejskie podejście do doskonałości i zaufania, 19.02.2020, Bruksela, [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_pl.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_pl.pdf) [dostęp: 1.04.2020].
- Cellan-Jones R., *Stephen Hawking warns artificial intelligence could end mankind*, 2.12.2014, BBC News, <https://www.bbc.com/news/technology-30290540> [dostęp: 1.04.2020].
- CEPEJ, *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment*. Adopted at the 31<sup>st</sup> plenary meeting of the CEPEJ (Strasbourg, 3–4 December 2018), Council of Europe 2019, <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c> [dostęp: 1.04.2020].
- Crawford K., *Artificial Intelligence's White Guy Problem*, „The New York Times”, 25.06.2016, <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html> [dostęp: 1.04.2020].
- Ethics Guidelines For Trustworthy AI*, 8.04.2019, Digital Single Market – European Commission, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> [dostęp: 1.04.2020].
- G20 Ministerial Statement on Trade and Digital Economy, <https://www.mofa.go.jp/files/000486596.pdf> [dostęp: 1.04.2020].
- Global Trendometer 2019: Essays on medium- and long-term global trends*, Global Trends Unit of the Directorate for the Library and Knowledge Services, Panel for the Future of Science and Technology, European Parliamentary Research Service, December 2019, [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/646111/EPRS\\_STU\(2019\)646111\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/646111/EPRS_STU(2019)646111_EN.pdf) [dostęp: 1.04.2020].

- Grodecka M., *A więc tak wygląda bunt maszyn. „Popelnij samobójstwo dla dobra planety” – taką poradę dał asystent głosowy*, 23.12.2019, Spider’s Web, <https://spidersweb.pl/2019/12/amazon-alexa-proponuje-samobojstwo.html> [dostęp: 1.04.2020].
- Hayes P.J., Morgenstern L., *On John McCarthy’s 80<sup>th</sup> Birthday, in Honor of His Contribution*, „AI Magazine” 2007, Vol. 28, no. 4, s. 93–102, [http://www-formal.stanford.edu/leora/hayes\\_morgenstern\\_birthday\\_mccarthy.pdf](http://www-formal.stanford.edu/leora/hayes_morgenstern_birthday_mccarthy.pdf) [dostęp: 1.04.2020].
- Preliminary study on the ethics of artificial intelligence*, World Commission on the Ethics of Scientific Knowledge and Technology, UNESCO 2019, <https://en.unesco.org/artificial-intelligence/ethics> [dostęp: 1.04.2020].
- Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, adopted by the Committee of Ministers on 8 April 2020 at the 1373<sup>rd</sup> meeting of the Ministers’ Deputies, <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016809e1154> [dostęp: 1.04.2020].
- Recommendation of the Council on Artificial Intelligence, adopted by the Council at Ministerial Level on 22 May 2019, C/MIN(2019)3/FINAL, OECD Publishing, [https://one.oecd.org/document/C/MIN\(2019\)3/FINAL/en/pdf](https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf) [dostęp: 1.04.2020].
- Report of COMEST on Robotics Ethics, SHS/YES/COMEST-10/17/2 REV, Paris, 14 September 2017, <https://unesdoc.unesco.org/ark:/48223/pf0000253952> [dostęp: 1.04.2020].
- de Strel A., Bibal A., Frenay B., Lognoul M., *Explaining the black box: when law controls AI*, 2020, [https://cerre.eu/wp-content/uploads/2020/03/issue\\_paper\\_explaining\\_the\\_black\\_box\\_when\\_law\\_controls\\_ai.pdf](https://cerre.eu/wp-content/uploads/2020/03/issue_paper_explaining_the_black_box_when_law_controls_ai.pdf) [dostęp: 1.04.2020].
- Sztuczna inteligencja dla Europy*, Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, COM(2018) 237 final, 25.04.2018, Bruksela.
- The ethics of artificial intelligence: Issues and initiatives*, Panel for the Future of Science and Technology, European Parliamentary Research Service 2020, [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS\\_STU\(2020\)634452\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf) [dostęp: 1.04.2020].
- Towards Responsible AI Innovation. Second INTERPOL-UNICRI Report on Artificial Intelligence for Law Enforcement*, 2020, [http://www.unicri.it/sites/default/files/2020-07/UNICRI-INTERPOL\\_Report\\_Towards\\_Responsible\\_AI\\_Innovation\\_0.pdf](http://www.unicri.it/sites/default/files/2020-07/UNICRI-INTERPOL_Report_Towards_Responsible_AI_Innovation_0.pdf) [dostęp: 1.04.2020].
- Understanding algorithmic decision-making: Opportunities and challenges*, Panel for the Future of Science and Technology, European Parliamentary Research Service 2019, [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS\\_STU\(2019\)624261\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf) [dostęp: 1.04.2020].
- Założenia do strategii AI w Polsce. Plan działań Ministerstwa Cyfryzacji, 9.11.2018, Warszawa, [https://www.gov.pl/documents/31305/436699/Za%C5%82o%C5%B-Cenia\\_do\\_strategii\\_AI\\_w\\_Polsce\\_-\\_raport.pdf](https://www.gov.pl/documents/31305/436699/Za%C5%82o%C5%B-Cenia_do_strategii_AI_w_Polsce_-_raport.pdf) [dostęp: 1.04.2020].
- Zou J., Schiebinger L., *AI can be sexist and racist — it’s time to make it fair*, 18.07.2018, nature, <https://www.nature.com/articles/d41586-018-05707-8> [dostęp: 1.04.2020].

**Abstract**  
**OECD Recommendations as an ethical and legal framework  
for the development of artificial intelligence**

New technologies using artificial intelligence play an increasingly important role in everyday life and are a source of profound social change. The prospect of the expansion of 'intelligent' machines that think and make decisions on their own is a source of concern not only for their users and manufacturers. It is necessary to determine and promote an ethical framework that guarantees respect for human values and interests by those developing technologies involving artificial intelligence. In the light of the rapidly increasing use of artificial intelligence in many sectors of the economy and its global reach, this is an extremely urgent issue. The article below presents the efforts made by the OECD to set global standards for developing artificial intelligence. The choice of this organisation is justified by the advanced level of work on guidelines for the development of artificial intelligence and the broad territorial coverage of the agreement between the international partners involved in this process. The OECD's recommendations on artificial intelligence have been accepted by 44 countries, including all leaders in the implementation of these technologies into the economy. They are therefore global in scope and provide a solid foundation for future regulation. The central pillar of the OECD recommendations are the principle of respect for human autonomy and dignity, human rights and democratic values. They promote artificial intelligence systems that can be considered trustworthy.

**Key words:** artificial intelligence, OECD recommendations, ethical principles, human rights

**Streszczenie**  
**Rekomendacje OECD jako ramy etyczne i prawne  
dla rozwoju sztucznej inteligencji**

Nowe technologie wykorzystujące sztuczną inteligencję odgrywają coraz poważniejszą rolę w życiu codziennym i są źródłem głębokich zmian społecznych. Perspektywa upowszechnienia „inteligentnych” maszyn, samodzielnie myślących i podejmujących decyzje, stanowi źródło niepokoju nie tylko dla ich użytkowników i producentów. Konieczne jest wyznaczenie i upowszechnienie ram etycznych gwarantujących poszanowanie ludzkich wartości i interesów przez podmioty rozwijające technologie wykorzystujące sztuczną inteligencję. W obliczu dynamicznie wzrastających zastosowań sztucznej inteligencji w wielu sektorach gospodarki oraz jej globalnego zasięgu, jest to niezwykle pilne zagadnienie. W poniższym tekście przedstawione zostały wysiłki podejmowane przez OECD w zakresie wyznaczenia globalnych standardów rozwoju sztucznej inteligencji. Wybór tej organizacji jest podyktowany zaawansowanym stopniem prac nad wytycznymi dotyczącymi rozwijania sztucznej inteligencji oraz szerokim zasięgiem terytorialnego porozumienia pomiędzy partnerami międzynarodowymi uczestniczącymi



w tym procesie. Rekomendacje OECD dotyczące sztucznej inteligencji zaakceptowały 44 państwa, w tym wszyscy liderzy we wdrażaniu tych technologii do gospodarki. Mają więc one zasięg globalny i stanowią solidny fundament dla budowania przyszłych regulacji prawnych. Elementem centralnym zaleceń OECD jest zasada poszanowania ludzkiej godności i godności, praw człowieka i wartości demokratycznych. Promują one systemy sztucznej inteligencji, które uznać można za godne zaufania.

**Słowa kluczowe:** sztuczna inteligencja, rekomendacje OECD, zasady etyczne, prawa człowieka